A Robust Approach to Credit Scoring with Deep Learning and Embedded Methods

Chinh Xuan Pham
Department of Information Technology, Banking Academy of Vietnam, Hanoi, Vietnam
chinhpx@hvnh.edu.vn

Huynh Ngoc Trinh
Institute for Artificial Intelligence, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
huynhtn@vnu.edu.vn

Long Quoc Tran
Institute for Artificial Intelligence, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
tqlong@vnu.edu.vn (corresponding author)

Received: 11 June 2025 | Revised: 18 September 2025 | Accepted: 27 September 2025

 DOI: https://doi.org/10.48084/etasr.12649

ABSTRACT

Credit scoring is essential for financial institutions to assess loan risk before making credit-granting decisions. Artificial Intelligence (AI) models are often applied to automate processes that support these organizations in decision-making. However, credit data is usually large and contains noisy or excessive features, which can degrade model performance and lead to inaccurate predictions. In this situation, feature selection is one of the most effective methods for improving model efficiency, as it identifies the most relevant attributes while reducing dimensionality and computational cost. This study proposes a robust pipeline that integrates an embedded feature selection method, either Lasso or Elastic Net, with deep learning models to enhance credit scoring performance. The proposed method was tested on five widely used financial datasets: the Credit Card database, the Australian Credit Approval dataset, the German Credit Data dataset, the Japanese Credit Screening dataset, and the Thomas Credit Risk dataset. The comparison results show that the proposed hybrid approach outperforms both the baseline methods and PCA-based feature selection in improving credit risk assessment.

Keywords-credit scoring; feature selection; deep learning; embedded methods

I. INTRODUCTION

Granting credit is an important activity in today's economy, as it allows individuals or businesses to secure a loan from a bank or other financial institution and commit to repaying it later, sometimes without interest. This, in turn, gives them more opportunities to expand their business and overcome crises, and indirectly stimulates economic development [1, 2]. However, credit is not always granted, as it should be based on assessment indicators regarding the level of creditworthiness and the payment ability of each borrower. In this scenario, credit scoring is a widely used method for evaluating consumers' credit profiles, ensuring prudent loan decisions. As a result, almost all financial transactions incorporate this procedure regardless of their scale or nature [3, 4]. Given the availability of such lending support, a substantial number of loan requests are generated. Banks and financial institutions must carefully evaluate these requests and make informed decisions to mitigate risks and maintain financial stability. To facilitate the process, many data-driven tools have been developed to assess loan applicants' creditworthiness and enhance the efficiency of credit decision-making [5].

With the advancement of technology, numerous tools and diverse methods have emerged, notably AI models. Some Machine Learning (ML) models have been effectively applied to the problem, such as logistic regression and decision trees [6]. In addition, advanced ML methods, such as random forests, which use ensemble techniques, yield improved results. ML is generally more effective than traditional methods that rely on simple statistics from a few basic criteria for this problem [7, 8]. However, this approach still faces limitations in capturing complex, non-linear patterns and handling large-scale, unstructured data. Deep Learning (DL) came later and emerged as a highly effective method. Its architectures can be stacked deeply to learn complex features, enabling the modeling of nonlinear relationships in large and diverse datasets [9]. In addition, there have also been attempts to apply deep generative models in capturing hidden data distributions and generating synthetic data for training [10].

In practice, processing large amounts of financial data often involves complex and redundant information that can adversely affect model performance. Many works have focused on developing feature selection techniques to identify and remove these uninformative characteristics [11, 12]. One of these is Fisher component-based methods, which use between-class and within-class variance to measure how well features can tell the difference between classes [13]. This gives a simple but useful way to rank them. Additionally, adaptive feature selection frameworks have also been proposed to automatically balance relevance and redundancy based on the data features, making the model work better [14].

First, this study proposes DL models to address the credit scoring problem. Then, embedded feature selection techniques, such as Lasso or Elastic Net, are integrated to eliminate redundant or irrelevant input data and enhance model performance. Five widely used public datasets are utilized in experiments applying seven standard evaluation metrics. Experimental results show that the proposed approaches, which integrate embedded feature selection into DL models, consistently outperform their counterparts without feature selection. These findings highlight the effectiveness and suitability of such integration for credit scoring tasks.

II. RESEARCH METHODOLOGY

This section presents the proposed method for the credit scoring problem. First, a comprehensive pipeline is proposed to apply DL models to this task, providing a methodical structure to grasp the data flow and maximize the effectiveness of the DL model. Then, two embedded methods are used for feature selection, Lasso and Elastic Net, to further enhance model performance. Their adoption is motivated by their ability to handle high-dimensional data and eliminate irrelevant or redundant features.

A. Pipeline

Figure 1 illustrates the proposed credit scoring pipeline, which is a general view of the process to develop feature selection methods with DL models for credit scoring. The pipeline starts by separating the unprocessed data into training and testing subsets, therefore guaranteeing an objective assessment of the performance of the model on fresh data. Next, it scales feature values to a uniform range using a scaler, especially MinMaxScaler. Particularly for distance-sensitive models, this stage helps the model run stably and increases performance and training speed. Feature selection with embedded techniques involves choosing pertinent data. This stage is crucial because it removes irrelevant data properties, aids in reducing data dimensionality, and enhances model performance and representation. The baseline scenario (None in the figure) represents the absence of feature selection, used for ablation study comparisons. Next, DL models are utilized, such as a Neural Network (NN) [15] or a Convolutional Neural Network (CNN) [16], trained on refined feature datasets. The final step evaluates the model's performance using seven metrics, including Accuracy (Acc), Precision (Prec), Recall (Rec), F1 score (F1), Area Under the Curve (AUC), Brier score (BS), and Kolmogorov-Smirnov (KS) statistic. The entire pipeline revolves around the combination of learning models with embedded feature selection methods to increase performance.

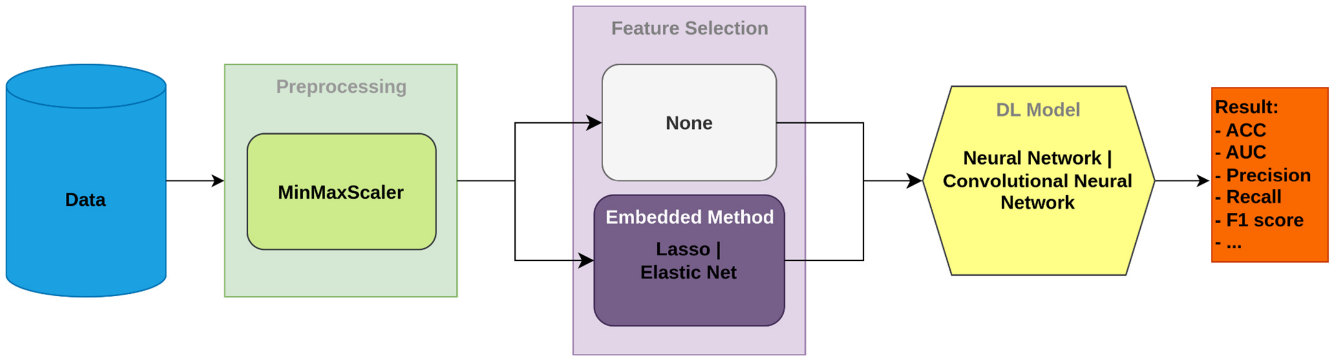

Fig. 1. Pipeline of credit scoring: From data collection to model evaluation.

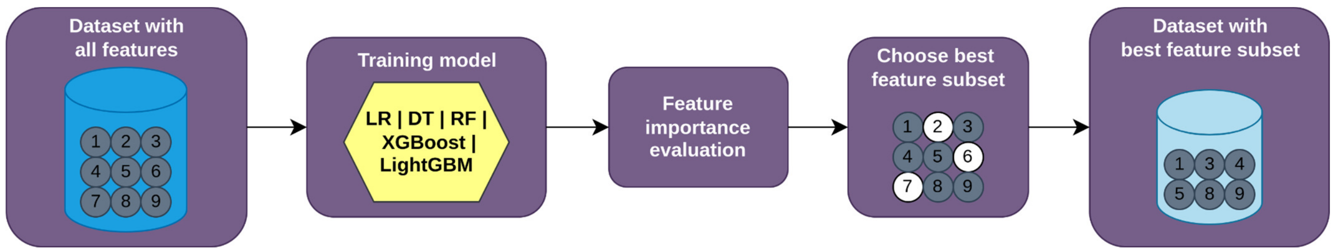

Fig. 2. Embedded methods pipeline.

B. Feature Selection

Embedded methods, as illustrated in Figure 2, incorporate feature selection into the model training process, enhancing both efficiency and predictive performance. This method performs feature selection as an integral part of the model training process. The procedure begins with the full set of features, followed by training multiple models and evaluating feature importance. During training, irrelevant or less important features are penalized or eliminated. This process is iterated until the optimal subset of features is identified. The selected features are then used by downstream models, improving generalization and robustness. This study focuses on embedded techniques derived from linear regression [17], where the model coefficients determine feature importance. These models optimize the predictions by minimizing the squared error between the actual and predicted values. Regularization techniques such as Lasso [18], Ridge [19], and Elastic Net [20], commonly applied in linear regression, help improve generalization and, in the case of Lasso and Elastic Net, directly eliminate redundant features.

1) Lasso

In Lasso regression [18], the coefficients are penalized by adding a constraint based on their absolute values. This regularization process helps reduce bias and prevent overfitting. The objective function for Lasso regression can be expressed as:

$$L(w) = \frac{1}{N} \times \|Xw - y\|_2{}^2 + a\|w\|_1 \qquad (1)$$

where X and y represent the input features and the output target for each data point, w denotes the model parameters, and $\|w\|_1$ is the L_1 norm of the coefficients, penalized by the regularization parameter a. What sets Lasso apart is its use of L_1 regularization, which can perform feature selection by reducing some coefficients to zero. This is due to the geometric nature of L_1 regularization. The penalty L_1 forms a diamond shape in a higher-dimensional space, with the corners of the diamond located along the coordinate axes. When optimizing the loss function, the optimal solution tends to fall in these sharp corners, where some coefficients are exactly zero. This results in a sparse solution vector, as irrelevant features are effectively eliminated. In contrast, Ridge regression applies the regularization L_2, which penalizes the squared values of the coefficients $\|w\|_2{}^2$ to shrink them. Unlike Lasso, it does not set the coefficients to zero but instead reduces their magnitudes. Ridge is effective for managing multicollinearity and is better suited for problems where feature selection is not the primary concern. Therefore, Lasso is especially effective for high-dimensional datasets where the selection of features is essential, whereas Ridge is better suited for managing multicollinearity without eliminating any features.

2) ElasticNet

Elastic Net [20] also extends logistic regression with regularization, enabling effective feature selection. This method utilizes both L_1 and L_2 penalties to leverage the advantages of both Lasso and Ridge regression. This method helps to overcome common limitations such as Lasso's tendency to select only one feature from a group of correlated variables and Ridge's inability to perform feature selection. The loss function of Elastic Net is updated as follows:

$$L(w) =$$
$$\frac{1}{N} \times \|Xw - y\|_2{}^2 + a(\frac{1-\rho}{2} \times \|w\|_2{}^2 + \beta\|w\|_1) \qquad (2)$$

where X is the input features, y is the corresponding target value, w denotes the model parameters, $\|w\|_1$ is the L_1 norm, and $\|w\|_2{}^2$ is the L_2 norm. The coefficient a in the formula controls the general strength of the regularization, in which β helps balance between the contributions of L_1 and L_2. Elastic Net is a relatively effective method that can outperform both Lasso and Ridge, especially on datasets exhibiting high multicollinearity or having more features than samples. In this scenario, Ridge may not eliminate irrelevant features, whereas Lasso may have difficulty consistently selecting a stable subset of features. This combination approach not only minimizes the coefficient values but also performs efficient feature selection. The Ridge's L_2 penalty discourages overly large coefficients, enhancing the model's stability and reducing the risk of overfitting. Meanwhile, the Lasso penalty encourages sparsity by pushing some coefficients exactly to zero, helping to isolate the most relevant features. These properties are particularly valuable in high-dimensional datasets and make Elastic Net a practical choice for complex tasks such as financial data analysis.

III. EXPERIMENTAL RESULTS

This section presents the evaluation of the proposed approach on five financial datasets related to the credit scoring problem. Initially, a summary of each dataset is provided to highlight its key features and differences. Next, the baseline performance of the DL models in addressing this task is presented. Scikit-learn was employed to implement an NN-based model with four hidden layers containing 10, 50, 100, and 200 neurons, respectively, each using the ReLU activation function, and a final sigmoid output layer for binary classification. For the CNN-based model, Keras was used to develop a one-dimensional CNN consisting of two Conv1D layers with ReLU activation, interleaved with Max-Pooling layers, followed by two fully connected layers and a sigmoid output layer. These architectures were selected after empirical evaluations, since deeper or more complex configurations led to overfitting without notable accuracy gains, while shallower structures failed to capture the nonlinear dependencies in the data. The chosen architectures provide a balanced trade-off between complexity, stability, and predictive performance. Finally, the effects of feature selection are examined using embedded techniques to enhance overall prediction performance and learning efficiency.

A. Dataset

All experiments were conducted on five benchmark credit scoring datasets obtained from public digital repositories, which provide a sufficiently large basis to enhance the robustness and credibility of this study. The Australian Credit Approval (Australian) dataset [21] is among the most frequently used in credit scoring research. It includes 14 input variables and 690 instances, representing credit card applications. Another widely adopted dataset is the German Credit Data (German) dataset [22], which classifies individuals as either good or bad credit risks based on a range of attributes. This study employed the alternative version containing only numerical attributes, resulting in 24 variables and 1000 instances. The Japanese Credit Screening (Japanese) dataset [23] was also used, which consists of 689 instances described by 15 variables. The above three datasets were obtained free of charge from the UCI Machine Learning Repository. The two remaining datasets were taken from books. The Credit Card Data (AER) dataset, available on the Kaggle platform, was originally published in [24]. It contains 1,319 instances and 11 input variables related to credit card applications. The final dataset, known as Thomas, was derived from [25]. Table I presents a summary of these datasets.

TABLE I. SUMMARY OF FIVE DATASETS

Dataset	No. of instances	Default /Not default	No. of features
Australian	690	383 / 307	14
German	1000	300 / 700	24
Japanese	689	383 / 306	15
AER	1319	296 / 1023	11
Thomas	1225	323 / 902	14

B. Baseline Results (without Feature Selection)

Table II presents the performance of the two base models in the five datasets. The results indicate that CNN generally outperforms NN in most cases, achieving higher values for Acc, AUC, F1, KS, Prec, and Rec, whereas maintaining a lower BS, suggesting improved predictive reliability. In particular, CNN demonstrates superior performance in three easier datasets, Australian, Japanese, and AER, where both models achieve relatively high Acc and AUC scores. The AER dataset yields the best overall performance among all cases, with the CNN model achieving 0.9500 in Acc and 0.9381 in AUC, generally outperforming the NN model across most metrics, except for Rec, where its performance is slightly lower. The CNN also demonstrates a better balance between Prec and Rec on the Japanese and Australian datasets, with Rec values reaching 0.8575 and 0.9608, respectively, indicating strong sensitivity to positive samples. The lower BS values in CNN indicate better probability calibration compared to NN across all three easier datasets, with the lowest value of 0.05 observed on the AER dataset and approximately 0.14 on the Australian and Japanese datasets. In contrast, performance deteriorates in the German and Thomas datasets, where both models exhibit lower F1 and AUC, indicating classification challenges. The Thomas dataset records the lowest predictive scores, with AUC values of approximately 0.52 for NN and 0.54 for CNN, highlighting difficulties in distinguishing between positive and negative samples. Rec values also remain significantly lower in German, indicating that both models struggle with minority class detection. The lower BS values in CNN still suggest better probability calibration, particularly in these two challenging datasets, where CNN continues to outperform NN despite the overall lower predictive performance. These findings suggest that, while CNN exhibits superior overall predictive performance, additional feature selection or model refinement is required for datasets such as German and Thomas, where classification remains challenging.

TABLE II. BASELINE RESULTS ON SEVEN METRICS

Dataset	Model	Metrics						
		Acc	AUC	F1	BS	KS	Prec	Rec
Australian	NN	0.8347	0.8344	0.8207	0.1653	0.6688	0.8103	0.8320
	CNN	0.8578	0.8583	0.8436	0.1422	0.7165	0.8321	0.8574
German	NN	0.7256	0.6658	0.5296	0.2744	0.3317	0.5482	0.5170
	CNN	0.7360	0.6773	0.5579	0.2640	0.3661	0.5820	0.5435
Japanese	NN	0.8524	0.8515	0.8308	0.1476	0.7030	0.8141	0.8488
	CNN	0.8561	0.8584	0.8469	0.1439	0.7169	0.8245	0.8734
AER	NN	0.9318	0.8871	0.9566	0.0682	0.7742	0.9462	0.9680
	CNN	0.9500	0.9381	0.9675	0.0500	0.8761	0.9753	0.9608
Thomas	NN	0.7296	0.5213	0.1312	0.2704	0.0433	0.4647	0.0780
	CNN	0.7287	0.5403	0.1964	0.2713	0.0807	0.5273	0.1261

C. Results with Feature Selection Based on Embedded Methods

Experiments were conducted using two embedded feature selection methods, namely Lasso and Elastic Net. In addition, the performance improvements were compared against the baseline and PCA-based feature selection, with 90% and 95% of the variance retained. These experiments provide a comprehensive analysis that demonstrates the effectiveness of the proposed method.

In general, most metrics achieved significant improvements when integrated with the embedded method, outperforming the baseline results (Table II). As shown in Table III, the integration of Lasso demonstrated considerable improvements in most evaluation metrics. In particular, notable enhancements in Prec, Rec, F1, and BS were shown in all datasets, exceeding the baseline results. The increase in F1 suggests that Lasso achieves a good trade-off between Prec and Rec, contributing to better classification performance. For example, on the Japanese dataset, Lasso combined with NN achieved an F1 of 0.8665, surpassing the baseline of 0.8308 by 3.57%. Similarly, on the Thomas dataset, this combination improved Rec from 0.4647 to 0.6284 and Prec from 0.0780 to 0.1234 with NN, highlighting its ability to capture relevant features more effectively. The reduction in BS, notably from 0.1653 to 0.1329 on the Australian dataset and from 0.0682 to 0.0436 on the AER dataset, reflects the role of Lasso in enhancing the quality of probabilistic prediction.

TABLE III. LASSO RESULTS ON SEVEN METRICS

Dataset	Model	Metrics						
		Acc	AUC	F1	BS	KS	Prec	Rec
Australian	NN	0.8671	0.8667	0.8504	0.1329	0.7333	0.8276	0.8749
	CNN	0.8624	0.8631	0.8483	0.1376	0.7263	0.8346	0.8640
German	NN	0.7462	0.6788	0.5439	0.2538	0.3575	0.5941	0.5065
	CNN	0.7352	0.6775	0.5385	0.2648	0.3551	0.5606	0.5290
Japanese	NN	0.8756	0.8792	0.8665	0.1244	0.7585	0.8222	0.9165
	CNN	0.8622	0.8620	0.8604	0.1378	0.7239	0.8444	0.8786
AER	NN	0.9564	0.9461	0.9714	0.0436	0.8922	0.9777	0.9654
	CNN	0.9485	0.9261	0.9673	0.0515	0.8522	0.9700	0.9647
Thomas	NN	0.7340	0.5448	0.2026	0.2814	0.0896	0.6284	0.1234
	CNN	0.7505	0.5466	0.2139	0.2495	0.0931	0.5342	0.1361

Furthermore, the Prec and Rec improvements, such as a Prec of 0.9700 and a Rec of 0.9647 on the AER dataset with CNN, underscore its effectiveness in maintaining predictive reliability. In addition, integrating Elastic Net with the DL models resulted in consistent improvements in various metrics, as shown in Table IV, compared to the baseline results. For example, on the Australian dataset, the CNN model achieved an F1 of 0.8438, reflecting a minor increase of 0.02% compared to the baseline. Similarly, on the German dataset, CNN with Elastic Net improved Acc by 0.64%, reaching 0.7424. On the Japanese dataset, the NN model achieved a Rec improvement of 4.88%, highlighting its increased sensitivity to predictions. Furthermore, Elastic Net consistently reduced the BS, reflecting improved probability calibration. This effect was most evident on the AER dataset, where BS decreased by 0.39% with the CNN model.

TABLE IV. ELASTIC NET RESULTS ON SEVEN METRICS

Dataset	*Model*	*Metrics*						
		Acc	**AUC**	**F1**	**BS**	**KS**	**Prec**	**Rec**
Australian	NN	0.8598	0.8592	0.8433	0.1402	0.7183	0.8462	0.8436
	CNN	0.8584	0.8579	0.8438	0.1416	0.7157	0.8453	0.8444
German	NN	0.7328	0.6723	0.5447	0.2672	0.3445	0.5867	0.5092
	CNN	0.7424	0.6817	0.5536	0.2576	0.3634	0.6004	0.5188
Japanese	NN	0.8646	0.8692	0.8462	0.1354	0.7383	0.8007	0.8976
	CNN	0.8634	0.8692	0.8483	0.1366	0.7384	0.8043	0.9006
AER	NN	0.9564	0.9430	0.9717	0.0436	0.8860	0.9755	0.9680
	CNN	0.9539	0.9534	0.9698	0.0461	0.9068	0.9849	0.9557
Thomas	NN	0.7368	0.5412	0.1960	0.2632	0.0824	0.5215	0.1244
	CNN	0.7290	0.5332	0.1875	0.2710	0.0665	0.4531	0.1233

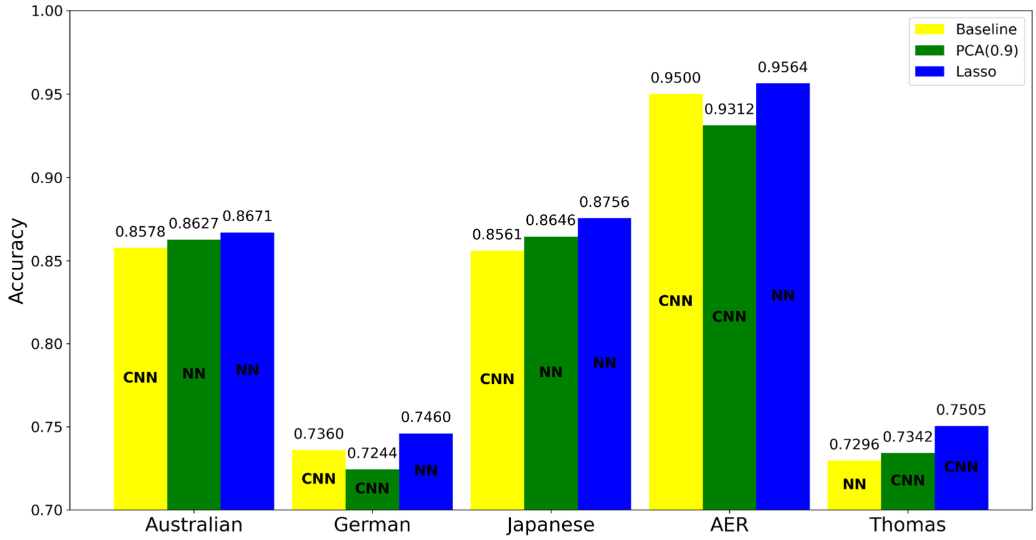

Fig. 3. Accuracy comparison among Baseline, PCA (0.9), and Lasso. The better-performing method between NN and CNN is chosen for plotting.

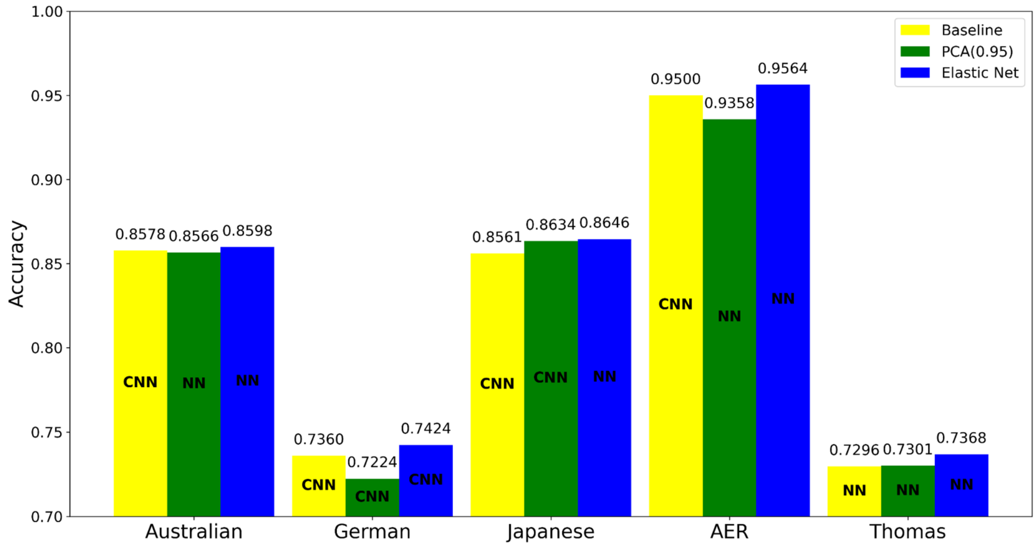

Fig. 4. Accuracy comparison between Baseline, PCA(0.95), and Elastic Net. The better-performing method between NN and CNN is chosen for plotting.

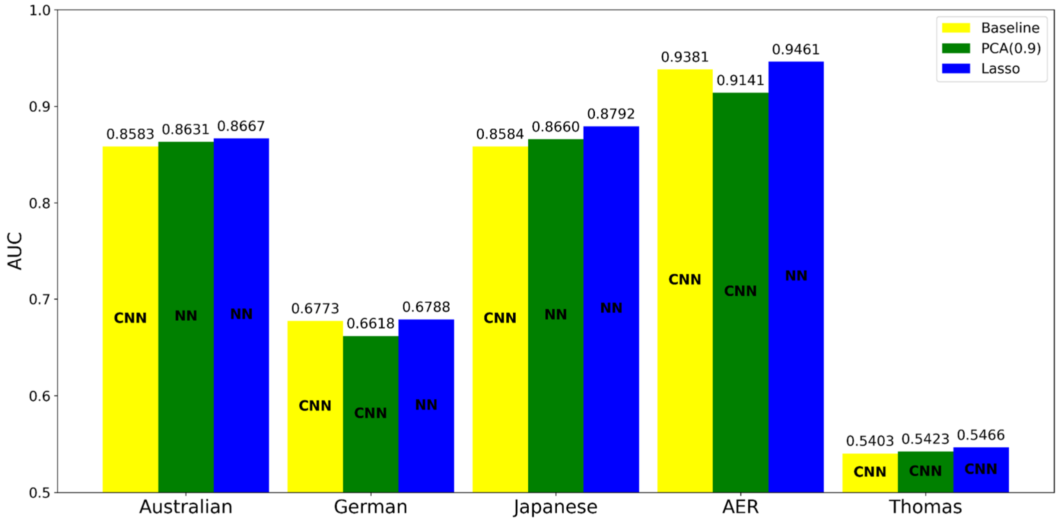

Fig. 5. AUC comparison between Baseline, PCA(0.9), and Lasso. The better-performing method between NN and CNN is chosen for plotting.

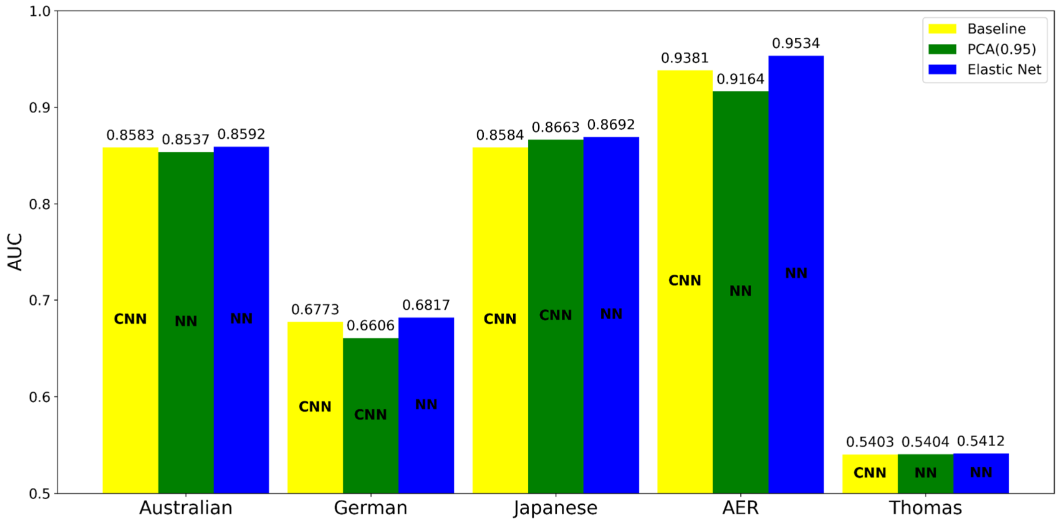

Fig. 6. AUC comparison between Baseline, PCA(0.95), and Elastic Net. The better-performing method between NN and CNN is chosen for plotting.

To extend the experiments for a comprehensive evaluation, the embedded feature selection methods were compared with PCA-based approaches. The best-performing CNN and NN models across datasets are shown in the bar charts in Figures 3 and 4 for Acc, and in Figures 5 and 6 for AUC, comparing PCA(0.90) and PCA(0.95). Overall, embedded methods consistently outperform both the baseline and PCA, with Lasso demonstrating the most stable and substantial improvements. As presented in Figures 3 and 4, both embedded methods achieve the highest performance compared to baseline and PCA, whether using PCA(0.9) or PCA(0.95). Lasso consistently improves accuracy by 1–2%, with the largest gain on the Thomas dataset being 1.6–2% higher than both the baseline and PCA(0.9). ElasticNet also shows improvements on multiple datasets, notably on Japanese and AER, with increases of 0.6–2% over the baseline and PCA(0.95).

In contrast, PCA-based feature selection exhibits inconsistent improvements, and in some cases, such as German and AER, accuracy is even lower than the baseline, highlighting its instability. The AUC results, shown in Figures 5 and 6, further confirm the superiority of the embedded methods. Lasso consistently improves AUC across all datasets, with gains ranging from approximately 0.5% to 1.5%, demonstrating high stability. ElasticNet provides notable increases on the AER dataset (from 0.9381 to 0.9534), while improvements on other datasets are moderate. PCA-based feature selection remains inconsistent and underperforms on German and AER datasets, where AUC values are even lower than the baseline, indicating its limited effectiveness.

Overall, these results indicate that both Lasso and ElasticNet consistently enhance nearly all evaluation metrics compared to the baseline and PCA-based feature selection methods. Lasso generally demonstrates superior performance across most datasets. For instance, on the Japanese dataset, Lasso achieves an accuracy of 0.8756 and an AUC of 0.8792, reflecting its strongest performance, while on the Australian dataset, it reaches an accuracy of 0.8671 and an AUC of 0.8667, further highlighting its effectiveness. In addition, ElasticNet remains competitive, but PCA-based feature selection shows inconsistent gains and underperforms on several datasets, such as German and AER. These findings indicate that Lasso is the most reliable and effective choice among the feature selection methods evaluated.

IV. CONCLUSIONS

This study investigates the potential of DL models, particularly NN and CNN, for credit scoring tasks. By integrating embedded feature selection methods, specifically Lasso and Elastic Net, the approach in this study significantly improved predictive performance and achieved greater stability across various datasets. Both methods contributed to enhanced results, with Lasso often delivering higher accuracy and demonstrating reliable performance on most datasets, although not consistently in every case, highlighting its practical applicability. Some main findings can be summarized as follows.

- A comprehensive workflow for the credit scoring problem that includes feature selection techniques to find irrelevant features, thus improving the performance of subsequent classification models.
- A methodical process for addressing feature selection in credit scoring using embedding techniques, including Lasso and Elastic Net.
- Comparative analysis between several datasets, metrics, and some methods to fully illustrate the efficacy of the approach, which is based on a combination of embedded methods and DL for feature selection in the credit scoring problem.

These results highlight the need to integrate embedded techniques with DL for feature selection to handle credit scoring problems. Future research could investigate either evaluating more advanced feature selection techniques or enhancing the integration of feature selection methods with new models based on transformer architectures. In addition, improving data preparation methods, such as resolving outliers or managing class imbalances, could help to increase model resilience and generalization for financial prediction activities.

REFERENCES

[1] K. Svirydzenka, "Introducing a New Broad-based Index of Financial Development," *IMF Working Papers*, vol. 16, no. 05, 2016, https://doi.org/10.5089/9781513583709.001.

[2] J. Ha, M. A. Kose, and F. Ohnsorge, *Inflation in Emerging and Developing Economies: Evolution, Drivers, and Policies*. World Bank Publications, 2019.

[3] M. A. Faheem, "AI-Driven Risk Assessment Models: Revolutionizing Credit Scoring and Default Prediction." Unpublished, 2021, https://doi.org/10.13140/RG.2.2.21281.01128.

[4] R. Hlongwane, K. K. K. M. Ramaboa, and W. Mongwe, "Enhancing credit scoring accuracy with a comprehensive evaluation of alternative data," *PLOS ONE*, vol. 19, no. 5, May 2024, Art. no. e0303566, https://doi.org/10.1371/journal.pone.0303566.

[5] S. Bi and W. Bao, "Innovative Application of Artificial Intelligence Technology in Bank Credit Risk Management," *International Journal of Global Economics and Management*, vol. 2, no. 3, pp. 76–81, Apr. 2024, https://doi.org/10.62051/IJGEM.v2n3.08.

[6] E. Dumitrescu, S. Hué, C. Hurlin, and S. Tokpavi, "Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1178–1192, Mar. 2022, https://doi.org/10.1016/j.ejor.2021.06.053.

[7] S. Mestiri, "Credit scoring using machine learning and deep Learning-Based models," *Data Science in Finance and Economics*, vol. 4, no. 2, pp. 236–248, 2024, https://doi.org/10.3934/DSFE.2024009.

[8] K. Tong, Z. Han, Y. Shen, Y. Long, and Y. Wei, "An Integrated Machine Learning and Deep Learning Framework for Credit Card Approval Prediction," in *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)*, Shenyang, China, July 2024, pp. 853–858, https://doi.org/10.1109/ICPICS62053.2024.10795883.

[9] Y. E. Gür, M. Toğaçar, and B. Solak, "Integration of CNN Models and Machine Learning Methods in Credit Score Classification: 2D Image Transformation and Feature Extraction," *Computational Economics*, vol. 65, no. 5, pp. 2991–3035, May 2025, https://doi.org/10.1007/s10614-025-10893-5.

[10] B. Alshawi, "Utilizing GANs for Credit Card Fraud Detection: A Comparison of Supervised Learning Algorithms," *Engineering, Technology & Applied Science Research*, vol. 13, no. 6, pp. 12264–12270, Dec. 2023, https://doi.org/10.48084/etasr.6434.

[11] J. Laborda and S. Ryoo, "Feature Selection in a Credit Scoring Model," *Mathematics*, vol. 9, no. 7, Mar. 2021, Art. no. 746, https://doi.org/10.3390/math9070746.

[12] O. Koc, O. Ugur, and A. S. Kestel, "The Impact of Feature Selection and Transformation on Machine Learning Methods in Determining the Credit Scoring." arXiv, 2023, https://doi.org/10.48550/arXiv.2303.05427.

[13] A. B. Buriro and S. Kumar, "The Fisher Component-based Feature Selection Method," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 9023–9027, Aug. 2022, https://doi.org/10.48084/etasr.5137.

[14] Y. El Touati, J. B. Slimane, and T. Saidani, "Adaptive Method for Feature Selection in the Machine Learning Context," *Engineering, Technology & Applied Science Research*, vol. 14, no. 3, pp. 14295–14300, June 2024, https://doi.org/10.48084/etasr.7401.

[15] M. C. Popescu, V. E. Balas, L. Perescu-Popescu, and N. Mastorakis, "Multilayer Perceptron and Neural Networks," *WSEAS Transactions on Circuits and Systems*, vol. 8, no. 7, pp. 579–588, July 2009.

[16] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks." arXiv, Dec. 02, 2015, https://doi.org/10.48550/arXiv.1511.08458.

[17] C. Thrampoulidis, S. Oymak, and B. Hassibi, "Regularized Linear Regression: A Precise Analysis of the Estimation Error," in *Proceedings of the 28th Conference on Learning Theory*, June 2015, pp. 1683–1709.

[18] V. Fonti and E. Belitser, *Feature Selection using LASSO*. VU Amsterdam, 2017.

[19] G. C. McDonald, "Ridge regression," *WIREs Computational Statistics*, vol. 1, no. 1, pp. 93–100, July 2009, https://doi.org/10.1002/wics.14.

[20] H. Zou and T. Hastie, "Regularization and Variable Selection Via the Elastic Net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 2005, https://doi.org/10.1111/j.1467-9868.2005.00503.x.

[21] R. Quinlan, "Statlog (Australian Credit Approval)." UCI Machine Learning Repository, 1987, https://doi.org/10.24432/C59012.

[22] H. Hofmann, "Statlog (German Credit Data)." UCI Machine Learning Repository, 1994, https://doi.org/10.24432/C5NC77.

[23] C. Sano, "Japanese Credit Screening." UCI Machine Learning Repository, 1992, https://doi.org/10.24432/C5259N.

[24] B. H. Baltagi, *Econometric Analysis of Panel Data*. John Wiley & Sons, 2005.

[25] J. Crook, "Credit Scoring and its Applications," *Journal of the Operational Research Society*, vol. 52, pp. 997–1006, 2002.